

Alternate Assessments in New Mexico

Responding to Community Recommendations

Learning Alliance New Mexico

Contents

Executive Summary.....	2
Overview	3
History of Assessment – What’s Shifted During the Past 15 Years?.....	4
How Does Assessment Impact Teaching and Learning?.....	6
How Does Assessment Impact Accountability?	10
Assessments in Other States.....	12
New Hampshire.....	12
California.....	14
Colorado.....	16
Local Opportunities for Federal Innovative Assessment Pilot.....	18
Existing Promising Practices.....	18
Conclusion.....	20
Appendix A – 34 Problems with Standardized Tests	22

Executive Summary

Assessments, like any other component of public education, can become more responsive – to students, educators and systems. The Every Student Succeeds Act, ESSA, provides opportunities for states to remodel their systems for greater alignment with local needs. Assessments can be a key design of this remodel.

Following up on the 2017 New Mexico Legislative Session Senate Joint Memorial 1, this report provides some information about assessment for the Legislative Education Study Committee to consider. After outlining a brief history of assessment during the past 15 years, how assessment impacts teaching and learning and how assessment impacts accountability, the report references three examples of how other states (New Hampshire, California and Colorado) assess learning. The final section before the conclusion underscores local opportunities for shifts in the New Mexican system and lifts up a sampling of existing promising practices in the state.

The report does not offer a comprehensive analysis of any of the sections described above. Rather, the report offers a glimpse into what has happened, what is happening and what could happen, which could be helpful context as the LESC reviews “No Time to Lose,” how other nations educate their citizens and how New Mexico could implement some of the global lessons learned.

In many other states, developing alternate assessments* has required years of large teams sharing their best thinking. While many conversations, meetings and reports focused on alternate assessments in New Mexico, a substantial amount of complicated work remains to create a more complete model. Fortunately, many New Mexicans are ready for that challenge.

**Note: While the term “alternate assessment” is used throughout the report to differentiate these models from “standardized assessment,” many stakeholders prefer the term “authentic assessment,” “innovative assessment,” or something more descriptive because “alternate” can carry a connotation of “less than.”*

Overview

The Every Student Succeeds Act, ESSA, passed with bi-partisan support during December 2015. ESSA replaced No Child Left Behind as the most recent iteration of the Elementary and Secondary Education Act, ESEA. ESSA encourages more state and local control of public education systems.

As such, ESSA required stakeholder engagement to help inform the state education agency, SEA, plan submitted to the federal Department of Education. Learning Alliance New Mexico, Learning Alliance, offered a decentralized process to engage stakeholders - resulting in more than 350 focus groups and feedback from 4,000 New Mexican participants.

When asked about high-quality student academic assessments and what strategies teachers could use to best gauge students' learning, participants advocated for:

Limited testing

Hands-on work

Individualized assessment

Participants suggested alternate ways students can demonstrate mastery of the knowledge, skills and attitudes that are important for success:

Projects

Collaboration

Application to Real Life

Self-Assessment

Verbal Communication

Additionally, participants recommended **assessment reforms** as a support for struggling schools¹.

During the 2017 New Mexico Legislation Session, legislators passed Senate Joint Memorial 1 “requesting the Public Education Department and the Legislative Education Study Committee to convene a working group to develop recommendations for a new student assessment policy aligned with the alternate assessment models allowed by the federal Every Student Succeeds Act.”² Learning Alliance offered to support the working group and efforts leading up to its convening. Row 29 of the 2017 LESC Interim Work Plan lists “Staff SJM 1 work group in developing recommendations for a new student assessment policy aligned with the alternative assessment models allowed by ESSA” and lists Learning Alliance as one of the Joint Project parties.³

This report serves as research to inform the LESC and future working groups.

¹ University of New Mexico Center for Education Policy Research. (March 2017). ESSA Phase II Stakeholder Feedback Report. Pp 1 – 2.

² *Senate Joint Memorial 1*. Retrieved from <http://www.sos.state.nm.us/uploads/files/SJM1-2017.pdf>

³ *LESC 2017 Interim Work Plan Summary*. Retrieved from https://www.nmlegis.gov/Publications/Work_Plans/ALESCworkplan17.pdf

History of Assessment – What’s Shifted During the Past 15 Years?

“At their core, educational assessments are statements about what educators, state policymakers, and parents want their students to learn and — in a larger sense — become...What we choose to assess is what will end up being the focus of classroom instruction. It is critical that the tests best represent the kind of learning students will need to thrive in the world that awaits them beyond graduation”⁴

Enacted in 2002, No Child Left Behind increased accountability by requiring testing of all students in grades 3 – 8 and some grades in high school. Results from the mandated tests could be used to sanction schools. During 2009, the federal Race to the Top program added a layer of complexity to the high-stakes testing by requiring part of teacher evaluations to include student achievement, as measured by test scores.⁵

High-stakes tests determine students’ readiness for school, track students into instructional levels, diagnose learning disabilities, and inform grade promotion or retention and graduation. Additionally, the tests inform curriculum content and instruction.⁶

During 2013, New Mexico adopted the Common Core State Standards, developed by the National Governors Association Center for Best Practices (NGA Center) and the Council of Chief State School Officers (CCSSO). The NGA Center and CCSSO “recognized the value of consistent, real-world learning goals and launched this effort to ensure all students, regardless of where they live, are graduating high school prepared for college, career and life.”⁷ Standards were informed by:

1. The best state standards already in existence
2. The experience of teachers, content experts, states and leading thinkers
3. Feedback from the public⁸

⁴ Gordon Commission on the Future of Assessment in Education. (2013). *A Public Policy Statement*. Retrieved from http://www.gordoncommission.org/rsc/pdfs/gordon_commission_public_policy_report.pdf

⁵ Nelson, Howard. (2013). *Testing More, Teaching Less*. Retrieved from <http://www.aft.org/sites/default/files/news/testingmore2013.pdf>

⁶ Fairtest. (2007, August 28). *How Standardized Testing Damages Education (Updated July 2012)*. Retrieved from <http://fairtest.org/how-standardized-testing-damages-education-pdf>

⁷ *Development Process*. Retrieved from <http://www.corestandards.org/about-the-standards/development-process/>

⁸ *Development Process*. Retrieved from <http://www.corestandards.org/about-the-standards/development-process/>

With new standards came a hope for new assessments. Fairtest.org published a list of Myths and Realities associated with the adoption of Common Core Standards, including those below:

<p>Myth: Common Core tests will be much better than current exams, with many items measuring higher-order skills.</p>	<ul style="list-style-type: none"> • Reality: <i>New tests will largely consist of the same old, multiple-choice questions.</i>
<p>Myth: Adoption of Common Core exams will end NCLB testing overkill.</p>	<ul style="list-style-type: none"> • Reality: <i>Under Common Core, there will be many more tests and the same misuses.</i>
<p>Myth: New multi-state assessments will save taxpayers money</p>	<ul style="list-style-type: none"> • Reality: <i>Test costs will increase for most states. Schools will spend even more for computer infrastructure upgrades.</i>
<p>Myth: New assessment consortia will replace error-prone test manufacturers.</p>	<ul style="list-style-type: none"> • Reality: <i>The same, incompetent, profit-driven companies will make new exams and prep materials.</i>
<p>Myth: More rigor means more, or better, learning.</p>	<ul style="list-style-type: none"> • Reality: <i>Harder tests do not make kids smarter.</i>
<p>Myth: Common Core assessments are designed to meet the needs of all students.</p>	<ul style="list-style-type: none"> • Reality: <i>The new tests put students with disabilities and English language learners at risk.</i>

9

During the same year, 2013, the Gordon Commission on the Future of Assessment in Education provided the following context:

<h2 style="margin: 0;">Historically</h2> <div style="background-color: #4F81BD; color: white; padding: 15px; border-radius: 10px; margin-top: 20px;"> <p>Educational assessments have been far more narrowly focused. Assessments have been designed primarily to provide summative information about student, teacher, school and system performance. That information has been used to highlight weaknesses, direct the spending of money, choose students for additional help or advanced classes, and evaluate the effectiveness of programs or teaching methods.</p> </div>	<h2 style="margin: 0;">Presently (2013)</h2> <div style="background-color: #76B82A; color: white; padding: 15px; border-radius: 10px; margin-top: 20px;"> <p>The federal government is absorbing the lion's share of the costs for the systems of assessment being developed by the PARCC and Smarter Balanced consortia. The conditions of that support stipulate that accountability components be the primary focus of their work. As a result, it is highly likely that the tools and resources needed to support teacher uses of assessment in the classroom will be seriously underdeveloped and in need of significant further work. When this round of federal funding ends, and the states are left with the challenges and costs associated with implementation and further development of accountability systems, there may be little money remaining to devote to formative assessment and practices.</p> </div>
--	--

⁹ Fairtest. (2013, September 3). *Common Core Assessment Myths and Realities: Moratorium Needed from More Tests, Costs, Stress*. Retrieved from <http://fairtest.org/common-core-assessments-factsheet>

New Mexico submitted a state plan to the federal Department of Education that carries forward emphasis on high-stakes standardized testing, specifically the Partnership of Readiness for College and Career, PARCC¹⁰. PARCC is “a group of states working together to develop a set of assessments that measure whether students are on track to be successful in college and careers.”¹¹ Twenty-four states and the District of Columbia began using PARCC when it was first released. As of October 2017, only six states and the District of Columbia continue to use PARCC¹². Many of the states in the initial consortium have opted out of the consortium, perhaps an important trend for New Mexico to consider why other states removed themselves.

How Does Assessment Impact Teaching and Learning?

“We know that annual assessments, as required by the ESSA, are tools for learning and promoting equity when they are done well and thoughtfully. When assessments are done poorly, in excess, or without a clear purpose, they take time away from teaching and learning.”¹³

“Throughout the long history of educational assessment in the United States, it has been seen by policymakers as a means of enforcing accountability for the performance of teachers and schools. For a relatively low outlay, assessments could expose academic weaknesses and make it possible to pressure schools and teachers to improve. But, as long as that remains their primary purpose, assessments will never fully realize their potential to guide and inform teaching and learning.”¹⁴

In many ways, assessment drives teaching and learning. With student, teacher and school accountability heavily influenced by assessment outcomes (see next section “How Does Assessment Impact Accountability?”), teaching and learning often focus on what’s assessed. Some may conclude this interconnectivity leads to deeper alignment between standards, teaching, learning and assessments. Others may conclude assessments pull teaching and learning in a direction less useful for real-life outcomes and more useful for test scores. If one considers the current assessment system as a pull-system (teaching and learning are pulled by the assessment), a provocative notion could be what would a push-system look like (where teaching and learning push into the assessment)?

Proponents of existing tests may assert the assessments enhance teaching and learning. On its website, PARCC references a string of 2015-2016 studies that “show PARCC is a high quality assessment, aligns to state learning standards, predicts college readiness, compares well to NAEP performance, is endorsed by the country’s top educators and gives students with disabilities and English learners more tools and access to the test than previous tests.”¹⁵

¹⁰ New Mexico Public Education Department. (2017, August 9). *New Mexico Rising – New Mexico’s State Plan for the Every Student Succeeds Act*. Retrieved from

http://www.ped.state.nm.us/ped/ESSA_docs/FINAL_NMESSAPlan.pdf

¹¹ PARCC. *Home*. Retrieved from <https://parcc.pearson.com/>

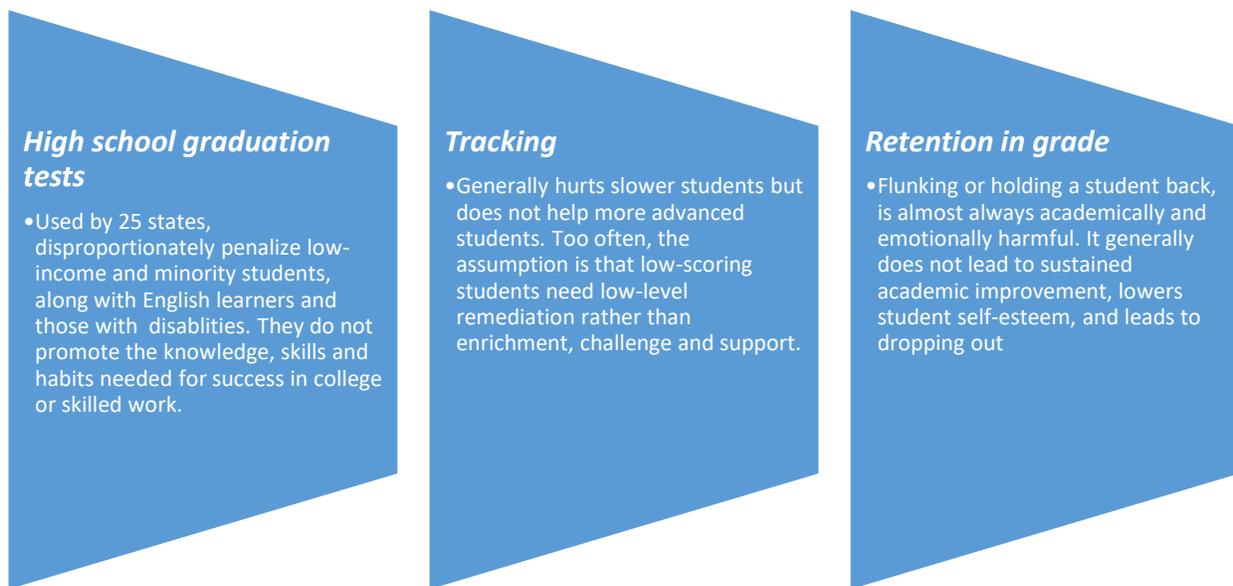
¹² Wikipedia. (2017, September 7). *PARCC*. Retrieved from <https://en.wikipedia.org/wiki/PARCC>

¹³ Department of Education. (2016, Dec. 8). *Federal Register*. Vol. 81 34 CFR Part 200 [Docket ID ED-2016-OESE-0047] RIN 1810-AB31

¹⁴ Gordon Commission on the Future of Assessment in Education. (2013). *A Public Policy Statement*. Retrieved from http://www.gordoncommission.org/rsc/pdfs/gordon_commission_public_policy_report.pdf

¹⁵ PARCC. *Independent Studies*. Retrieved from <http://parcc-assessment.org/assessments/test-design/independent-studies>

Critics of the existing tests offer a different narrative. Participants of the Learning Alliance ESSA Stakeholder Engagement focus groups said PARCC is the least valuable of the tests offered in school¹⁶. Other issues include:



Furthermore, unintended consequences of **screening** and **readiness** tests are frequently inaccurate results that lead to misdiagnosis of student learning needs – disproportionately adversely impacting students from low-income and minority-group backgrounds, English learners, and students with disabilities, who are more likely to be denied diplomas, retained in grade, placed in a lower track, or unnecessarily put in remedial education programs. The students mentioned above are more likely to receive less challenging curriculum, based heavily on rote drill and test practice, which influences the students to fall further and further behind their peers. Many drop out, some ending up in the “school-to-prison pipeline.” On the other hand, children from white, middle and upper income backgrounds are more likely to be placed in “gifted and talented” or college preparatory programs where they are challenged to read, explore, investigate, think and progress rapidly.¹⁷

For a review of a sample of other issues, The Washington Post published an article April 19, 2017 titled “34 Problems with Standardized Tests.” That list can be found in Appendix A.

When New Mexicans consider how they want assessments to impact teaching and learning, they can look to local organizations that worked with local people to identify the kind of learning students will need to thrive in the world. For example, Mission: Graduate, a collective impact backbone organization housed in the United Way Central New Mexico, developed a “Graduate Profile” highlighting the knowledge, skills and attitudes students will need to succeed in life. How do current assessments ensure the positive characteristics identified in the “Graduate Profile?” What would a future assessment look like to more accurately encourage and evaluate the same characteristics?

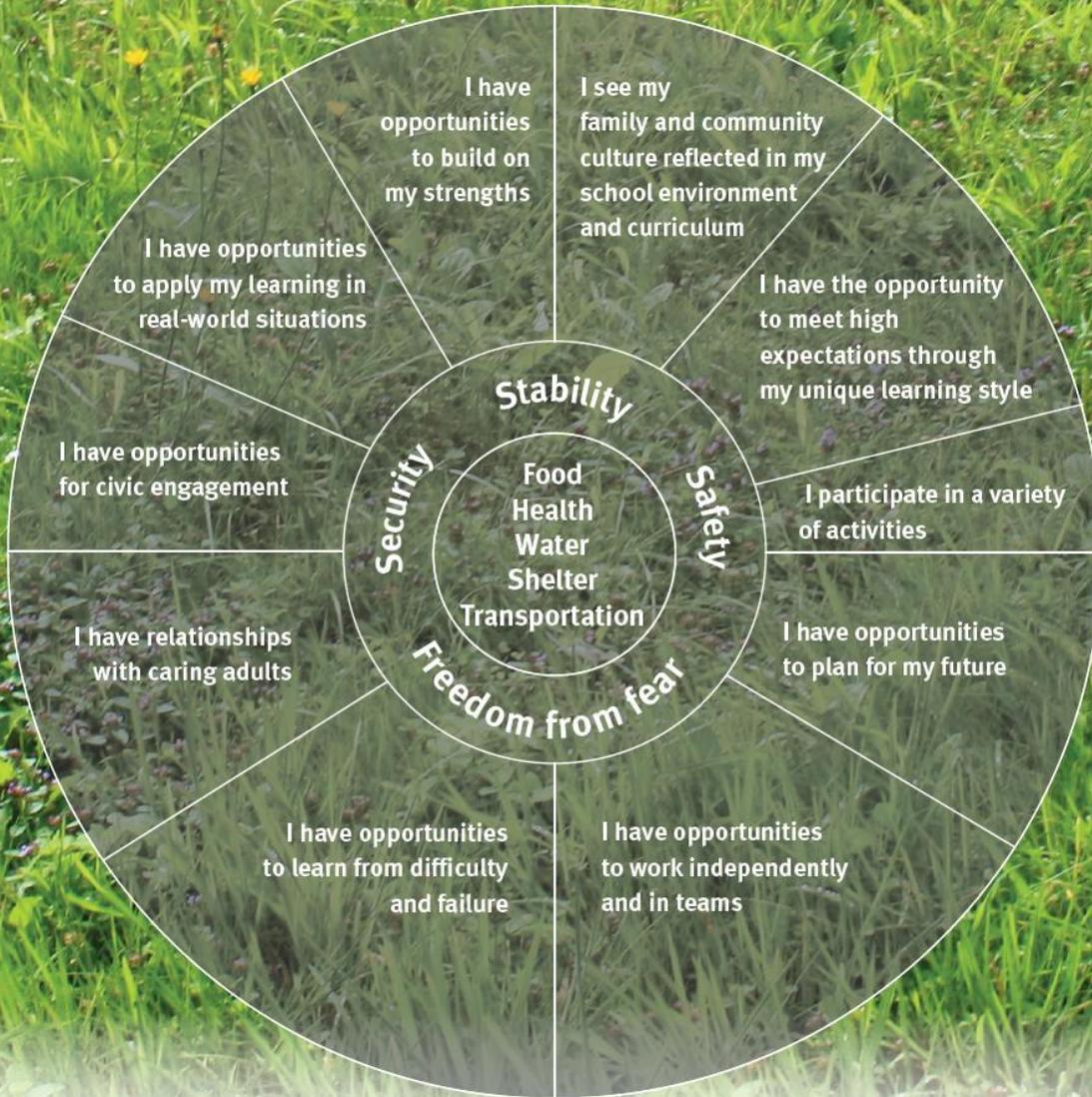
¹⁶ University of New Mexico Center for Education Policy Research. (March 2017). *ESSA Phase II Stakeholder Feedback Report*. P 9.

¹⁷ Fairtest. (2007, August 28). *How Standardized Testing Damages Education (Updated July 2012)*. Retrieved from <http://fairtest.org/how-standardized-testing-damages-education-pdf>

CENTRAL NEW MEXICO GRADUATE PROFILE



THE LAUNCHPAD: Foundational Experiences



Community voices informed the development of the Graduate Profile and Launchpad. Mission: Graduate, an education partnership, facilitated 51 focus groups with principals and teachers, postsecondary faculty, employees from different sectors, young people, parents, and community members. More information is available in the Graduate Profile User Guide.



18

¹⁸ Mission: Graduate. *Graduate Profile*. Retrieved from <https://missiongraduatnm.org/graduate-profile>

Some argue the current assessment framework prioritizes efficiency over effectiveness. Important outcomes, like those listed in the Graduate Profile, may be more difficult to measure. It may be easier to implement an assessment that measures that which is easier to measure. Whatever the structure of the assessment, teachers will be responsible to reverse-engineer their teaching to drive toward what is measured. When teachers steer instruction toward test prep “drill and kill” – practice questions that mimic the test – they steer away from students preferred learning methods of engaging projects, hands on learning and discussions.¹⁹

“Policy makers and researchers should design and evaluate new test-based incentive programs in ways that provide information about alternate approaches to incentives and accountability. This should include exploration of the effects of key features suggested by basic research, such as who is targeted for incentives; what performance measures are used; what consequences are attached to the performance measures and how frequently they are used; what additional support and options are provided to schools, teachers and students in their efforts to improve; and how incentives are framed and communicated. Choices among the options for some or all of these features are likely to be critical in determining which – if any- incentive programs are successful.”

20

How Does Assessment Impact Accountability?

“In many districts, standardized exam results have become the single most important indicator of school performance. As a result, teachers and administrators feel enormous pressure to ensure that test scores consistently rise. Schools narrow and manipulate the curriculum to match the test, while teachers tend to cover only what is likely to be on the next exam. Methods of teaching conform to the multiple-choice format. Education increasingly resembles test prep. It is easy to see why this could happen in low-scoring districts. But some high-scoring schools and districts, striving to keep their top rank, also succumb. The pressure is so great that a growing number of administrators and teachers have engaged in various kinds of cheating to boost scores.”²¹

“Completely eliminating all testing is both unrealistic and undesirable. However, cutting testing time and costs in half would yield significant gains both to the instructional day and to the budget.”²²

Assessments are intricately connected to accountability. Assessments serve as the foundation for student, teacher and school accountability systems. In New Mexico, test results inform students’ ability to graduate, teachers’ performance evaluations and school report cards.²³

¹⁹ Fairtest. (2007, August 17). *Multiple Choice Tests*. Retrieved from <http://www.fairtest.org/facts/mctfcats.html>

²⁰ National Research Council. (2011). "Summary." *Incentives and Test-Based Accountability in Education*. Washington, DC: The National Academies Press. doi: 10.17226/12521

²¹ Fairtest. (2007, August 28). *How Standardized Testing Damages Education (Updated July 2012)*. Retrieved from <http://fairtest.org/how-standardized-testing-damages-education-pdf>

²² Nelson, Howard. (2013). *Testing More, Teaching Less*. Retrieved from <http://www.aft.org/sites/default/files/news/testingmore2013.pdf>

However, many say assessment results cannot reliably, validly or fairly be used to judge students, teachers or schools. Popular value-added methods, VAMs, of teacher evaluation have been found full of errors. The effects to morale – of students, parents, teachers, administrators – can be harmful. If standardized testing is the primary means of accountability, public education becomes “accountable” to the testing industry.²⁴ ***What needs to shift so public education becomes primarily accountable to the local communities served?***

How can New Mexico build a comprehensive assessment system that uses standardized assessment as only a component? Fairtest.com recommends better methods of evaluating students such as:

- Careful observation and documentation of student work and behaviors by trained teachers who have inter-rater reliability (similar to the judges who score Olympic athletes)
- Assessment based on student performance on real learning tasks²⁵

New Mexicans identified additional approaches such as:

Projects	Self-Assessment	Collaboration
Real-life application		Verbal communication

26

New Mexicans seem to be asking for what other nations already use: performance-based assessments such as essays, projects and activities to evaluate students. Many other nations score higher on international exams because of their focus on performance-based approaches, instead of focusing on multiple choice and short answer tests, as the United States does.²⁷

ESSA encourages states to pursue Innovative Assessment Pilots that explore options how assessments can be used.

“Accurate and reliable measurement of student achievement based on annual State assessments in reading/language arts and mathematics remains a core component of State assessment and accountability systems under the ESSA. While States are required to use the results of State assessments in statewide accountability systems, consistent with sections 1111(c) and 1111(d) of the ESEA, there are no further requirements for how assessment results are used, including for teacher evaluation or student advancement and promotion decisions.”

28

²³ New Mexico Public Education Department. (2017, August 9). *New Mexico Rising – New Mexico’s State Plan for the Every Student Succeeds Act*. Retrieved from http://www.ped.state.nm.us/ped/ESSA_docs/FINAL_NMESSAPlan.pdf

²⁴ Fairtest. (2007, August 28). *How Standardized Testing Damages Education (Updated July 2012)*. Retrieved from <http://fairtest.org/how-standardized-testing-damages-education-pdf>

²⁵ Fairtest. (2007, August 28). *How Standardized Testing Damages Education (Updated July 2012)*. Retrieved from <http://fairtest.org/how-standardized-testing-damages-education-pdf>

²⁶ University of New Mexico Center for Education Policy Research. (March 2017). *ESSA Phase II Stakeholder Feedback Report*. P. 2.

²⁷ Fairtest. (2007, August 28). *How Standardized Testing Damages Education (Updated July 2012)*. Retrieved from <http://fairtest.org/how-standardized-testing-damages-education-pdf>

²⁸ Department of Education. (2016, Dec. 8). *Federal Register*. Vol. 81 34 CFR Part 200 [Docket ID ED–2016–OESE–0047] RIN 1810–AB31

Assessments in Other States

“While the Department [of Education] appreciates the need to allow States flexibility in designing innovative assessments, this flexibility must be balanced with the imperative that States meet all of the statutory provisions and ensure their innovative assessment systems are valid, reliable, fair, and of high-quality.”²⁹

States approach assessment differently. This section provides information about a sampling of states that may be of interest to New Mexico as New Mexico considers alternate assessments. One of the most important states to review is New Hampshire as ESSA includes the “Assessment and Accountability Demonstration Authority” which was modeled after New Hampshire’s PACE initiative and extends this opportunity for up to seven states.

New Hampshire

From the New Hampshire Department of Education – “New Hampshire was awarded permission from the U.S. Department of Education in March 2015 to pilot an accountability system designed to support deeper learning for students and powerful organization change for schools and districts. The accountability pilot, referred to as Performance Assessment of Competency Education or PACE, is grounded in a competency-based educational approach designed to ensure that students have meaningful opportunities to achieve critical knowledge and skills. There are several key components to this approach:

- Explicit involvement of local educational leaders in designing and implementing the accountability system,
- Intense and reciprocal support on behalf of the NH DOE for local districts involved in this initiative that will include technical, policy, and practical guidance,
- Use of competency-based approaches to instruction, learning, and assessment which can best support the goal of college and career readiness for all students, and
- Use of instructionally-relevant, high-quality performance-based assessments, alongside periodic administration of Smarter Balanced and SAT assessments of state standards in math and English language arts (ELA), for the purpose of tracking and reporting the progress of students, schools, districts, and educators.³⁰

The foundational principles of this initiative are:

- Provide students with real opportunities to learn the knowledge, skills, and work-study practices necessary for all students to graduate from high school college-and-career ready, pursue meaningful post-secondary options and to become productive citizens of NH and the world
- Promote and measure knowledge, skills, and dispositions in ways that provide feedback for improvement of student and system learning

²⁹ Department of Education. (2016, Dec. 8). *Federal Register*. Vol. 81 34 CFR Part 200 [Docket ID ED-2016-OESE-0047] RIN 1810-AB31

³⁰ New Hampshire Department of Education. (2016, January). *Moving from Good to Great in New Hampshire: Performance Assessment of Competency Education (PACE)*. Retrieved from <https://www.education.nh.gov/assessment-systems/documents/overview.pdf>

- Performance assessments are a vehicle for encouraging the teaching and learning of meaningful content and skills
- Build the capacity to engage students in the learning of meaningful knowledge and skills
- Accountability systems, including educator evaluation systems, should be built upon a foundation of a student assessment system that measures critically important student outcomes
- Common performance tasks have high technical quality
- Locally designed performance tasks with guidelines for ensuring high technical quality
- Regional scoring sessions and local district peer review audits to ensure sound accountability systems and high inter-rater reliability
- Web-based bank of local and common performance tasks
- Regional support network for districts and schools.³¹

The New Hampshire PACE testing schedule looks like:

Grade	ELA	Math	Science
K-2	Local Performance Assessments	Local Performance Assessments	Local Performance Assessments
3	Smarter Balanced & Local PBAs	Common and Local PBAs	Local Performance Assessments
4	Common and Local PBAs	Smarter Balanced & Local PBAs	Local Performance Assessments
5	Common and Local PBAs	Common and Local PBAs	Common PBAs
6	Smarter Balanced & Local PBAs	Common and Local PBAs	Local Performance Assessments
7	Common and Local PBAs	Smarter Balanced & Local PBAs	Common PBAs
8	Common and Local PBAs	Common and Local PBAs	Local Performance Assessments
9	Interim Smarter Balanced & Local PBAs	Common and Local PBAs	Local Performance Assessments
10	Common and Local PBAs	Interim Smarter Balanced & Local PBAs	Common PBAs
11/12	Smarter Balanced (plus optional SAT)	Smarter Balanced (plus optional SAT)	Local Performance Assessments
	CAPSTONE PERFORMANCE ASSESSMENT		

32

³¹ New Hampshire Department of Education. *NH Performance Assessment Network*. Retrieved from <https://www.education.nh.gov/assessment-systems/>

³² New Hampshire Department of Education. *New Hampshire Accountability Pilot Overview – Performance Assessment of Competency Education (PACE)*. Retrieved from <https://www.education.nh.gov/assessment-systems/documents/pilot-overview.pdf>

California

From the California Department of Education – “The California Assessment of Student Performance and Progress (CAASPP) System offers:

- *Smarter Balanced Summative Assessments for English language arts/literacy (ELA) and mathematics in grades three through eight and eleven.*
 - All students at the designated grade levels are required to participate except:
 - Students who participate in the alternate assessments
 - ELA only – English learners who are in their first 12 months of attending a school in the United States

Content and Format: The Smarter Balanced Summative Assessments which are delivered by computer consist of two sections: a computer-adaptive test and a Performance Task (PT) based on the Common Core State Standards for ELA and mathematics. The computer-adaptive section includes a range of item types such as selected response, constructed response, table, fill-in, graphing, etc. The PT are extended activities that measure a student’s ability to integrate knowledge and skills across multiple standards – a key component of college and career readiness.

- *California Alternate Assessments (CAAs) for ELA and mathematics in grades three through eight and grade eleven.*
 - Students with the most significant cognitive disabilities who are unable to take the Smarter Balanced Summative Assessments even with accessibility supports and whose individualized education program (IEP) indicates assessment with an alternate test

Content and Format: The content of the alternate assessment are based on alternate achievement standards derived from the Common Core State Standards for students with the most significant cognitive disabilities. The CAAs are a computer-based two-stage adaptive test. The CAAs are administered to students in a one-on-one environment with a testing examiner who is familiar with the student.

- *California Science Test (CAST) Field Test in grades five, eight and assigned grade in high school*
 - The CAST is required from all students in grades, five, eight and assigned grade in high school unless their IEP indicates assessment with an alternate test. The CAST includes a stand-alone or discrete items and performance tasks (PTs). The discrete item types consist of selected response, constructed response, table, fill-in, graphing, etc. The PTs are extended activities that measure a student’s ability to integrate knowledge and skills across multiple standards.

Content and Format: The CAST is aligned with the California Next Generation Science Standards (CA NGSS). The CA NGSS were adopted by the State Board of Education September 2013. The 2017 administration of the CAST was a census pilot test and the 2018 CAST will be a census field test.

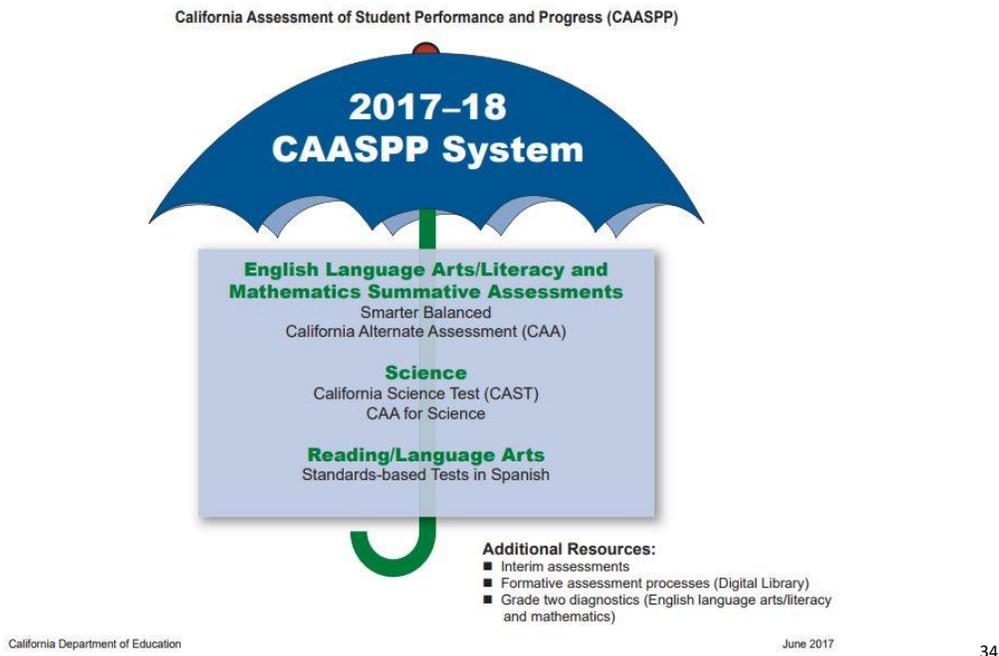
- *California Alternate Assessment (CAA) for Science Pilot Test in grades five, eight and assigned grade in high school*

- Students with the most significant cognitive disabilities who are unable to take the CAST even with accessibility supports and whose IEP indicates assessment with an alternate test. The CAA for Science is administered to students in grades five, eight and assigned grade in high school.

Content and Format: The content of the alternate assessment are based on alternate achievement standards derived from the CA NGSS for students with the most significant cognitive disabilities. Test examiners administer the CAA for Science students in a one-on-one environment. Pilot year one the CAA for Science consisted of one embedded performance task and the second year pilot, consists of three embedded performance tasks that can administered once to students over the course of the winter/spring.

- *Standards-based Tests in Spanish (STS) for reading/language arts in grades two through eleven (optional)*
 - The STS is a primary language test in Spanish for Spanish-speaking English learners who either receive instruction in their primary language or have been enrolled in a school in the United States less than 12 months. The STS is administered at the discretion of the local educational agency. Therefore, it is considered an optional assessment.

Content and Format: The STS is aligned to the California content standards for reading/language arts. The State Board of Education adopted blueprints for this test by grades and over time between 2005 and 2007. For the 2017-2018 administration, the STS for reading/language arts will be administered on the computer and is comprised entirely of multiple-choice questions.”³³



³³ California Department of Education. (2017, July 31). *CAASPP Description – CalEdFacts*. Retrieved from <http://www.cde.ca.gov/ta/tg/ai/cefcaaspp.asp>

³⁴ (2017, June). *CAASPP*. Retrieved from <http://www.cde.ca.gov/ta/tg/ca/documents/caaspp16umbrella.pdf>

Colorado

From the Colorado Department of Education – “In 2009 Colorado implemented more rigorous academic standards to ensure that all Colorado students are prepared for college and career success. Fully implemented in the 2013-14 school year, the Colorado Academic Standards set clear, consistent guidelines for what students should know and be able to do at each grade level across 10 subject areas, including English language arts, math, science and social studies. The rigorous standards were designed by educators and researchers, with significant input from higher education and workforce leaders, to help students prepare for life after high school. Prior to the adoption of the standards, too many students in Colorado were graduating from high school without the problem-solving and critical-thinking skills needed to succeed in college and in the workforce. These standards were designed to change that.

To accurately measure student mastery of these skills and expectations, Colorado adopted assessments that align with the Colorado Academic Standards. Colorado Measures of Academic Success (CMAS) assessments are the state’s common measurement tool of student progress at the end of the school year in English language arts, math, science and social studies. Designed to be administered online, the tests feature a variety of interactive questions that are engaging and aligned with 21st century teaching and learning practices. Across all 10 content areas, the standards demand more of our students and teachers than previous standards.

- *RIGOR* - Students use inquiry, critical thinking and creative processes to solve problems
- *RELEVANCY* - Students engage in real world scenarios that require the application and transfer of concepts and skills
- *DISCIPLINARY LITERACY* - Students learn to speak, think, work and argue like mathematicians, artists, scientists, authors and historians

The standards-aligned assessments help educators and parents know how their students are performing and growing over time and how they compare to their peers across the district and the state.”³⁵

Scoring CMAS Tests - Who scores the tests?

Qualified scorers are recruited from across the country and must have a Bachelor’s degree or higher in mathematics, English, education or a related field. Scorers for the mathematics portions of the CMAS assessments hold at least a four-year degree in a related field and have demonstrated the knowledge needed to effectively score responses to math questions. Scorers for the English language arts and literacy portions of the CMAS assessments hold at least a four-year degree in English, education, history, psychology, journalism or a related field, and/or teacher certification or other work experience that will enable them to effectively score the literacy analysis, research simulation, or narrative writing tasks found in the CMAS assessments. Not all scorers are teachers, but as many as three-quarters have previous teaching experience. Half of all CMAS scorers are current K-12 teachers.

How are scorers trained?

All scorers receive extensive training to evaluate student performance on a select and specific group of questions across multiple exams. To ensure that scoring is fair and unbiased, scorers are trained on specific questions instead of a student’s full assessment. Each scorer receives extensive

³⁵ Colorado Department of Education. (2017, August 14). *Colorado Measures of Academic Success: What to Expect for the 2017-2018 School Year*. Retrieved from <https://www.cde.state.co.us/communications/20170814whattoexpect>

training at a regional scoring center on his or her specific question or group of questions and must pass two evaluations before they are deemed eligible to score an assessment.

How are the tests scored?

There is a regimented and defined process to score every student exam that ensures accuracy and security of the assessments and students' information. First, all students are assigned an identification number to protect their privacy. Student answers are then separated and sorted question-by-question and sent to the scorers who have been trained and qualified to score that particular question. This maintains student anonymity and allows scorers to become experts in scoring one question at a time. Scorers assign points to each answer. Depending on the question, up to six points could be available. Each scorer has a binder for each question with the scoring rubric and examples of pre-scored answers that he or she can use to compare his or her scoring against the guide prepared by educators. To ensure scorers are maintaining accuracy standards throughout the scoring process, they will routinely be given pre-scored answers along with un-scored answers. A scorer's evaluations must match the "true" scores at least 70 percent of the time. When a scorer's accuracy declines, he or she receives additional training on the test question. If a scorer cannot maintain consistency and accuracy, his or her previous scores are all put back into the system for re-scoring.³⁶

Spring 2018 and 2019 Colorado State Assessment Tentative Windows				
Assessment	Content Area	Grades	Spring 2018 TENTATIVE Testing Windows	Spring 2019 TENTATIVE Testing Windows
ACCESS for ELLs®	English language proficiency	K-12	January 8 – February 9, 2018	January 14 – February 15, 2019
Colorado Measures of Academic Success (CMAS)	Mathematics	3 – 8	Official Window: April 9 – 27, 2018	Official Window: April 8 – 26, 2019
	English language arts (including CSLA)	3 – 8	Early HS Science/Social Studies Window: March 26 – April 13, 2018 OR April 2 – April 20, 2018	Early HS Science/Social Studies Window: March 25 – April 12, 2019 OR April 1 – April 19, 2019
	Science	5, 8, & 11		
	Social Studies**	4, 7, & 11	Extended Math/ELA/CSLA Window: Begin as early as March 19, 2018*	Extended Math/ELA/CSLA Window: Begin as early as March 18, 2019*
Colorado Alternate Assessment (CoAlt)	Mathematics (DLM)	3-11	Official Window: April 9 – 27, 2018	Official Window: April 8 – 26, 2019
	English language arts (DLM)	3 – 11	Early HS Science/Social Studies Window: March 26 – April 13, 2018 OR April 2 – April 20, 2018	Early HS Science/Social Studies Window: March 25 – April 12, 2019 OR April 1 – April 19, 2019
	Science	5, 8, & 11		
	Social Studies**	4, 7, & 11	Extended Math/ELA Window: Begin as early as March 19, 2018*	Extended Math/ELA Window: Begin as early as March 18, 2019*
CO PSAT	English language arts and Mathematics	9 & 10	District choice for initial test date: April 10, 11, or 12, 2018	District choice for initial test date: April 9, 10, or 11, 2019
			Accommodations window: April 10 – 17, 2018	Accommodations window: April 9 – 16, 2019
			Make-up testing window: April 11 – 20, 2018:	Make-up testing window: April 10 – 19, 2019
CO SAT	College Entrance Exam	11	Initial test date: April 10, 2018	Initial test date: April 9, 2019
			Accommodations window: April 10 – 13, 2018	Accommodations window: April 9 – 12, 2019
			Make-up test date: April 24, 2018	Make-up test date: April 23, 2019

37

³⁶ Colorado Department of Education. (2017, October 3). *Assessment Fact Sheet*. Retrieved from <http://www.cde.state.co.us/communications/factsheetsandfaqs-assessment>

³⁷ Colorado Department of Education. (2017, October 11). *General Assessment Information*. Retrieved from <https://www.cde.state.co.us/assessment/generalinfo>

Local Opportunities for Federal Innovative Assessment Pilot

“ESSA includes in Title I, Part B of the ESEA a new demonstration authority under which an SEA or consortium of SEAs that meets certain application requirements may establish, operate, and evaluate an innovative assessment system, including for use in the statewide accountability system, with the goal of using the innovative assessment system after the demonstration authority ends to meet the academic assessment and statewide accountability system requirements under Title I, Part A of the ESEA...Section 1204(e)(2)(A)(vi) of the ESEA requires that assessments be developed so that they are accessible to all students, including English learners and students with disabilities; are fair, valid, and reliable; and hold all students to the same high standards.”³⁸

The Department of Education offers demonstration authority to seven states. ESSA allows for several options how states develop, implement and scale an innovative assessment. States may partner with a district or school. States may choose to focus on a single grade and content area. However the state begins its work under the new demonstration authority, the intention will be to scale the innovative assessment statewide within five years.³⁹

New Mexico has existing innovative assessment practices to launch from if the state chooses to apply for new demonstration authority. Following up on findings from the Learning Alliance Report, the table below lifts up a sample of examples of the practices aligned with recommendations from ESSA Stakeholder Engagement participants:

Existing Promising Practices

Group	Assessment	Innovation
Albuquerque Charter Academy	Internally-Designed Portfolio	Components aligned with Common Core requirements
Architecture, Construction and Engineering Leadership High School	NM PAN Shell	Thorough, rigorous tracking of comprehensive work to ensure validity and reliability of performance assessments
Dual Language Education New Mexico	Bilingual Seal ⁴⁰	Case study: Albuquerque High School uses a portfolio assessment requiring demonstration of proficiency in language and literacy in Spanish/English and academic proficiency demonstrated in Spanish across content areas

³⁸ Department of Education. (2016, Dec. 8). *Federal Register*. Vol. 81 34 CFR Part 200 [Docket ID ED-2016-OESE-0047] RIN 1810-AB31

³⁹ Department of Education. (2016, Dec. 8). *Federal Register*. Vol. 81 34 CFR Part 200 [Docket ID ED-2016-OESE-0047] RIN 1810-AB31

⁴⁰ *APS Bilingual Seal Handbook*. Retrieved from <http://www.aps.edu/language-and-cultural-equity/aps-bilingual-seal/aps-bilingual-seal-handbook>

Group	Assessment	Innovation
Families United for Education	Social Justice Seal	Social justice practitioners certify students' skills after students pass an ethnic studies class with strong analysis of race and racism
Native American Community Academy	Wellness Wheel ⁴¹	Self-assessment of holistic wellbeing
New Mexico Center for School Leadership	New Metrics ⁴²	Performance-based assessment, comprehensive exhibition, community members evaluate
New Mexico School for the Arts	Iterative Assessment – Performance and Academic	Blends growth mindset and experiential learning with internally-created system
NM STEM Ecosystem	Various locally-designed assessments in individual programs	Science assessments informed by STEM experts, practitioners and employers

The New Mexico Center for School Leadership published a number of reports that provide additional context and highlight local practices and opportunities:

- **The New Metrics Final Report** – highlighting local work⁴³
- **New Metrics Comparative Analysis** – comparing and contrasting New Mexicans systems with practices in other states (New Hampshire & New York)⁴⁴
- **Student Voice Report** – students offer insights to improve the New Metrics Outcomes Tool⁴⁵

The Public Education Department convened a work group for alternate demonstration of competency during its ESSA Stakeholder Engagement work. That group encouraged alternative assessments (primarily portfolios) to have the following components:

- Standards alignment
- Common rubrics
- Locally-scored
- Existing work form as early as 9th grade
- Removal of standardized assessment retakes in state regulation⁴⁶

⁴¹ *Native American Community Academy Personal Learning Plan*. Retrieved from <http://nacaschool.org/wp-content/uploads/2010/06/Wellness-Wheel.pdf>

⁴² New Mexico Center for School Leadership. *New Metrics Outcomes Alignment Activities Notes*. Retrieved from <http://leadership-pdc.org/wp-content/uploads/2017/11/OutcomesAlignmentNotes.pdf>

⁴³ New Mexico Center for School Leadership. (2017). *New Metrics Final Report*. Retrieved from <http://leadership-pdc.org/wp-content/uploads/2017/11/New-Metrics-Final-Report-Final.pdf>

⁴⁴ New Mexico Center for School Leadership. (2017, February). *New Metrics: A Comparative Analysis of Local and National Assessment Practices*. Retrieved from <http://leadership-pdc.org/wp-content/uploads/2017/11/New-Metrics-Comparative-analysis-Final-1.pdf>

⁴⁵ New Mexico Center for School Leadership. (2016, May). *New Metrics Student Voice Report*. Retrieved from <http://leadership-pdc.org/wp-content/uploads/2017/11/Student-Voice-Engagement-Report-FINAL.pdf>

Blending components from local findings and practices could propel innovation assessment in New Mexico. Many community members demand innovation. Many practitioners offer innovation. How can the state build a responsive bridge?

Senate Joint Memorial 1 from the 53rd Legislature Session calls for a working group of “educational practitioners; academic and pedagogical researchers; community, tribal and civic leaders; union members; industry representatives; higher education representatives; and others who have an interest in developing an alternative assessment model to build upon the work that is already taking place in the state and to develop recommendations for a new assessment policy⁴⁷.” This report provides some research on which a robust working group could build.

In addition to further developing a local assessment - compiled from promising practices and responsive to community needs – the working group could borrow from recommendations of the Gordon Commission and:

- ✓ Conduct research on how assessments are changing and help inform NMPED so the department makes good purchasing decisions, and addresses issues as they arise.
- ✓ Mount a public education campaign targeting parents, educators, school board members and the media explaining the importance of good assessment to quality education.
- ✓ Study “equitable assessment” and explore issues related to diversity, equity and excellence.⁴⁸

Conclusion

Student assessment informs student learning. Student assessment informs classroom teaching. Student assessment informs teacher and school accountability. Student assessment informs public education systems. Many New Mexicans call for innovative assessment to improve student learning, classroom teaching, accountability and public education systems.

Some current criticisms of “teaching to the test” indicate teachers spend too much time preparing students to become proficient test-takers and not enough time preparing students to become successful, engaged citizens; “drill and kill” rote preparation impedes hands-on, project-based learning; focus solely on academics compromises focus on social/emotional learning, arts and physical education; outdated evaluations do not prepare students for the jobs of tomorrow. What if “teaching to the test” was an enriching endeavor?

Imagine if there was broad community agreement on a local assessment worthy of being taught to; one informed by local practitioners, employers and content-area experts. Imagine if the assessments generated deep learning, in addition to scoring it. Imagine if the local assessment provided data on how well New Mexico prepared its students to be successful New Mexicans, as defined by local “Graduate Profiles.” Possibilities extend as far as local imagination.

⁴⁶ New Mexico Public Education Department. *Appendix B: ESSA Technical Workgroup – Future Ready Students*. Retrieved from http://www.ped.state.nm.us/ped/ESSA_docs/04112017/ESSAAppendicesA-Y_jg_1.pdf

⁴⁷ *Senate Joint Memorial 1*. Retrieved from <http://www.sos.state.nm.us/uploads/files/SJM1-2017.pdf>

⁴⁸ Gordon Commission on the Future of Assessment in Education. (2013). *A Public Policy Statement*. Retrieved from http://www.gordoncommission.org/rsc/pdfs/gordon_commission_public_policy_report.pdf

DOE recognizes, *“The primary benefit of these regulations is the administration of statewide assessments that more effectively measure student mastery of challenging State academic standards and better inform classroom instruction and student supports, ultimately leading to improved academic outcomes for all students. We believe that this benefit outweighs associated costs to an SEA, which may use funds received under the Grants for State Assessments and Related Activities program and funds reserved for State administration under part A of title I to participate in the demonstration authority. In addition, high-quality, innovative assessment models developed by participating SEAs under the demonstration authority can benefit other SEAs by providing examples of new assessment strategies for those SEAs to consider.”*⁴⁹

Other states are finding ways to better serve students through innovative assessment. ***How will New Mexico create its own solutions?***

⁴⁹ Department of Education. (2016, Dec. 8). *Federal Register*. Vol. 81 34 CFR Part 200 [Docket ID ED–2016–OESE–0047] RIN 1810–AB31

Appendix A – 34 Problems with Standardized Tests

Commercially produced machine-scored standardized tests:

1. Are unavoidably biased by social-class, ethnic, regional, and other cultural differences.
2. Unfairly advantage those who can afford test prep.
3. Radically limit teacher ability to adapt to learner differences.
4. Provide minimal to no useful feedback to classroom teachers.
5. Are keyed to the deeply flawed, knowledge-fragmenting “core” curriculum adopted in 1893.
6. Have led to the neglect of play, music, art and other nonverbal ways of learning.
7. Hide problems created by margin-of-error computations in scoring.
8. Penalize test-takers who think in nonstandard ways (which the young frequently do).
9. Give control of the curriculum to test manufacturers.
10. Encourage use of threats, bribes, and other extrinsic motivators to raise scores.
11. Assume that what the young will need to know in the future is already known.
12. Emphasize minimum achievement to the neglect of maximum performance.
13. Produce scores which can be — and sometimes are — manipulated for political purposes.
14. Create unreasonable pressures to cheat.
15. Use arbitrary, subjectively-set pass-fail cut scores.
16. Reduce teacher creativity and the appeal of teaching as a profession.
17. Lessen concern for and use of continuous evaluation.
18. Have no “success in life” predictive power.
19. Unfairly channel instructional resources to learners at or near the pass-fail cut score.
20. Are open to scoring errors with life-changing consequences.
21. Are at odds with deep-seated American values about individuality and worth.
22. Create unnecessary stress and negative attitudes toward schooling.
23. Perpetuate the artificial compartmentalization of knowledge by field.
24. Channel increasing amounts of tax money away from classrooms and into corporate coffers.
25. Waste the vast, creative potential of human variability.
26. Block instructional innovations that can’t be evaluated by machine.
27. Unduly reward mere ability to retrieve secondhand information from memory.
28. Subtract from available instructional time.
29. Lend themselves to “gaming” — strategies to improve the success-rate of guessing.
30. Make time — a parameter largely unrelated to ability — a factor in scoring.
31. Create test fatigue, aversion, and eventual refusal to take tests seriously.
32. Hide poor quality test items behind secrecy walls.
33. Undermine a fundamental democratic principle that those closest to the work are best positioned to evaluate its quality.
34. According to the National Academy of Sciences report to Congress, don’t increase student achievement.

At the most fundamental level, education policy shaped by standardized test scores is at odds with the deepest of all societal needs — human survival. Inevitable environmental, demographic, technological, institutional, and cognitive system changes require continuous adaptation. Adaptation requires new knowledge. New knowledge is generated by dozens of complex thought processes — hypothesizing, inferring, relating, valuing, imagining, and so on. And of those dozens of complex thought processes, only two — recalling, and applying — can be quantified and measured with sufficient precision to produce a meaningful number.

50

⁵⁰ Strauss, Valerie. (2017, April 19). *34 Problems with Standardized Tests*. Retrieved from https://www.washingtonpost.com/news/answer-sheet/wp/2017/04/19/34-problems-with-standardized-tests/?utm_term=.7743ae7497a4&wpisrc=nl_answer&wpmm=1